P1 224518

THE UNITED STATES OF AMERICA

TO ALL TO WHOM THESE PRESENTS SHALL COME:

UNITED STATES DEPARTMENT OF COMMERCE
United States Patent and Trademark Office

March 29, 2000

THIS IS TO CERTIFY THAT ANNEXED HERETO IS A TRUE COPY FROM THE RECORDS OF THE UNITED STATES PATENT AND TRADEMARK OFFICE OF THOSE PAPERS OF THE BELOW IDENTIFIED PATENT APPLICATION THAT MET THE REQUIREMENTS TO BE GRANTED A FILING DATE UNDER 35 USC 111.

APPLICATION NUMBER: *60/120,801*
FILING DATE: *February 19, 1999*
PCT APPLICATION NUMBER: *PCT/US00/04331*

By Authority of the
COMMISSIONER OF PATENTS AND TRADEMARKS

P. R. GRANT

PRIORITY
DOCUMENT
SUBMITTED OR TRANSMITTED IN

'PROVISIONAL APPLICATION FOR PATENT COVER SHEET

This is a request for filing a PROVISIONAL APPLICATION FOR PATENT under 37 CFR §1.53(b)(2).

Docket No. 99,068	Type a plus sign (+) inside this box:	+

INVENTOR(S)/APPLICANTS(S)

LAST NAME	FIRST NAME	MIDDLE INITIAL	RESIDENCE (City and either state or foreign country)
Wang	Jian		Pittsburgh, PA
Harrington	Chris		Pittsburgh, PA
Taylor	Lans		Pittsburgh, PA

TITLE OF THE INVENTION (280 character maximum)

Database For Storage, Retrieval, and Analysis Of Cellular Information

CORRESPONDENCE ADDRESS

McDonnell Boehnen Hulbert & Berghoff
300 South Wacker Drive, Chicago

STATE	Illinois	ZIP CODE	60606	COUNTRY	U.S.A.

ENCLOSED APPLICATION PARTS (check all that apply)

X	Specification	Number of Pages 9	☐	Small Entity Statement
☒	Drawing(s)	Number of Sheets 3	☐	Other (specify):

METHOD OF PAYMENT FOR THIS PROVISIONAL APPLICATION FOR PATENT

XX	A check or money order is enclosed to cover the Provisional Filing Fee. The Commissioner is hereby authorized to charge filing fees and credit Deposit Account Number: 13-2490.	PROVISIONAL APPLICATION FOR PATENT FILING FEE AMOUNT ($)	75.00

The invention was made by an agency of the United States Government or under a contract with an agency of the United States Government.
_X_ No. _____ Yes, the name of the U.S. Government agency and the Government contract number are: _____

Respectfully submitted,
SIGNATURE: _____ Date: 2/19/99

TYPED or PRINTED NAME David Harper   REG. NO. 42,636

Additional inventors are being named on separately numbered sheets attached hereto.

USE ONLY FOR FILING A PROVISIONAL APPLICATION FOR PATENT

Field of the Invention

This invention relates to a system and database for the storage, retrieval, and analysis of cellular information.

5

Background of the Invention

Traditionally, cell biology research has largely been a manual, labor intensive activity. With the advent of tools that can automate much cell biology experimentation (see for example U.S. Patent Application SN 08/810,983 filed February 27, 1997, incorporated by reference herein in its entirety), the rate at which complex information is generated about the functioning of cells has increased dramatically. As a result, cell biology is not only an academic discipline, but also the new frontier for large-scale drug discovery. How to present, organize and analyze the complex information about cell functioning so that new knowledge can be generated is critical for both pharmaceutical research and basic cell biology research.

Current technology uses simple text-based presentation of cellular information, or presents only a subset of the information associated with a cellular entity or event. This invention enables the user to visually interact with all identified dimensions of cellular information at the same time and dynamically navigate through those dimensions to find out the relationship of one piece of information with other pieces of information. This process facilitates the abstraction of knowledge from information. The Cellomics™ Database uses pathways to capture cellular knowledge. The pathway knowledge is then used as a portal to unite all other cellular information, thus enabling the synthesis of new knowledge by investigating the inner relationship of this information.

Other related systems* only capture a small subject area while the Cellomics™ Database is an open system that integrates a wide variety of cellular information. Current technology uses simple text-based presentation of cellular information, or presents only a subset of the information associated with a cellular entity or event.

5   * Including
   1) Ecocyc from Pangea (Exemplified in Nucleic Acids Research 26:50-53 (1998); Ismb 2:203-211 (1994)
   2) KEGG pathway database from Institute for Chemical Research, Kyoto University (Nucleic Acids Research 27:377-379 (1999); Nucleic Acids Research 27:29-34
10   (1999))
   3) CSNDB from Japanese National Institute of Health Sciences (Pac Symp. Biocomput 187-197 (1997)
   4) SPAD from Graduate School of Genetic Resources Technology, Kyushu University, Japan
15   5) PUMA (http://www-c.mcs.anl.gov/home/compbio/PUMA/Production/Graphics/puma_graph.html) from Computational Biology in the Mathematics and Computer Science Division at Argonne National Laboratory.

20   The present invention enables the user to visually interact with all identified dimensions of cellular information at the same time and dynamically navigate through those dimensions to find out the relationship of one piece of information with other pieces of information. This process facilitates the abstraction of knowledge from information.

25

This user interface has the following key features:

1)   Dynamic generation of pathway diagrams to represent cellular functions (see Figure 1).

2)   The diagrams of feature 1 capture the spatial information about each entity in the
30   diagram by associating each entity with a specific cellular compartment.

3)   The diagram of feature 1 is used as a navigation tool to retrieve information associated with certain cellular functions or entities.   Information is presented hierarchically, from more general to more specific. Color-coding is used to reflect the highest level of generalization.

5   4)   Cellular information is organized into dimensions. Each dimension is organized into hierarchies of information.

5)   Every cellular entity has some information associated with it in each dimension as defined in feature 4.   When an entity is selected in the diagram of feature 1, its corresponding information in the dimensions as well as its position in the relevant

10   hierarchy is dynamically presented.

6)   Interactive updating of the diagrams of feature 1. Users can selectively expand and/or collapse parts of the diagram, or rearrange the layout of the diagram. Updating of the diagram can be achieved by making restrictions on some dimensional information, such as "only show the entities that have been shown to be functional in certain cell

15   types." Updated information can replace the old information or the old information and updated information can be presented in different planes using 3-dimensional diagrams or in different windows. When they are presented in different planes, the planes can be parallel to each other or at an angle (for example, 90 degrees).

7)   This visual presentation is also used as the basis for a cell editor to input

20   information about cellular pathways. The user will draw a cellular pathway onto predefined cell templates and a program will capture information about that pathway as it is drawn. When more information is necessary or when there is ambiguity, a software program will prompt the user for clarification. Before committing the input, a textual

description will be shown to the user so that the user can confirm that the computer has correctly captured his/her intentions while interacting with the graphical interface. The user can directly edit the textual information before it is submitted.

8)   As the user interacts with the database through various visualizations described above, a history file is kept to record his/her activities. Upon request, a graphical representation of these activities can be plotted.

9)   Different shapes are used to represent different types of entities in the diagram of feature 1.

10)   The user can define his/her own diagram from the underlying data in a database.

11)   The user can zoom and pan.

Cellomics™ Database uses a standard based, platform independent means to transmit information (XML) which enables the system to more easily integrate with other public domain or proprietary information source.

This invention can be used to facilitate the user's understanding of cell functioning, to design experiments more intelligently and to analyze experimental results more thoroughly. Specifically, this invention can help drug discovery scientists select better targets for pharmaceutical intervention in the hope of curing diseases.

This invention is a complete system that enables the easy storage, retrieval, and analysis of cellular information. Figure 2 shows the main system components of this invention, which includes the Cellomics™ Database itself, the application software that runs on the Cellomics™ Database servers and the client machines. Client machines can access the Cellomics™ Database through either the Internet or an Intranet. Cellomics™

Database application software will access proprietary databases within a customer site or public domain databases through the Internet. Figure 3 shows the general steps in interacting with this system. A user can interact with the system in either edit mode or query mode. Different users will be assigned privileges to either only query Cellomics™

5 Database or both edit and query the database. Appendix A shows data representation schema in the form of an XML DTD.

It should be understood that the programs, processes, methods and databases described herein are not related or limited to any particular type of computer or network

10 system (hardware or software), unless indicated otherwise. Various types of general purpose or specialized computer systems may be used with or perform operations in accordance with the teachings described herein.

In view of the wide variety of embodiments to which the principles of the present invention can be applied, it should be understood that the illustrated embodiments are

15 exemplary only, and should not be taken as limiting the scope of the present invention. For example, the steps of the flow diagrams may be taken in sequences other than those described, and more or fewer elements may be used in the block diagrams. While various elements of the preferred embodiments have been described as being implemented in software, in other embodiments hardware implementations may

20 alternatively be used and visa-versa.

```
                  A DTD for cellular pathway informaion: PML.dtd    -->
<!-- Author(s):  Jian Wang -->
<!-- Log:    Start: 12/15/98/JW       -->
<!-- Copyright: Cellomics, Inc-->

<!ELEMENT Pathways (Pathway*)>

<!-- ref defines all the references used in an xml doc.  reflink at this level
links to the
references that is generaic to the whole pathway.  reflink at other levels are
references
specific to that level -->
<!ELEMENT Pathway
((BioSys|Component|Cell_Compartment|Cellular_Process|Functional_Unit|Transformat
ions|FeatureInfo|Ref)*,RefLink?, Notes*)>
<!ATTLIST Pathway
            Pathway_ID ID #REQUIRED
            Pathway_Name CDATA #IMPLIED>


<!ELEMENT BioSys (Organism?,System?,Organ?,Tissue?,Cell?,Notes*)>
      <!ATTLIST BioSys
            BioSys_ID ID #REQUIRED>
      <!ELEMENT Organism EMPTY>
      <!ATTLIST Organism
            Organism CDATA #REQUIRED
            DevStage CDATA #IMPLIED>
      <!ELEMENT System EMPTY>
      <!ATTLIST System
            System CDATA #REQUIRED
            DevStage CDATA #IMPLIED>
      <!ELEMENT Organ EMPTY>
      <!ATTLIST Organ
            Organ CDATA #REQUIRED
            DevStage CDATA #IMPLIED>
      <!ELEMENT Tissue EMPTY>
      <!ATTLIST Tissue
            Tissue CDATA #REQUIRED
            DevStage CDATA #IMPLIED>
      <!ELEMENT Cell EMPTY>
      <!ATTLIST Cell
            Cell CDATA #REQUIRED
            CellCycleStage CDATA #IMPLIED
            DevStage CDATA #IMPLIED>


<!ELEMENT Cell_Compartment (#PCDATA|Notes)*>
      <!ATTLIST Cell_Compartment
            Compartment_ID ID #REQUIRED
            Compartment_Name CDATA #REQUIRED>


<!ELEMENT Cellular_Process (#PCDATA|Notes)*>
      <!ATTLIST Cellular_Process
            Process_ID ID #REQUIRED
            Process_Name CDATA #REQUIRED>
```

```
<!ATTLIST Component
        Component_ID ID #REQUIRED
        Component_Name CDATA #REQUIRED
        BioSys IDREF #REQUIRED>

<!ELEMENT Modification (#PCDATA|Notes)*>
      <!ATTLIST Modification
            Modification_Site CDATA #IMPLIED
            Modification_Type CDATA #REQUIRED>

<!ELEMENT Functional_Unit (ComponentLink*, Synonym*, RefLink?, Notes*)>
      <!ATTLIST Functional_Unit
            Unit_ID ID #REQUIRED
            Unit_Name CDATA #REQUIRED
            Unit_Abbr CDATA #IMPLIED
            BioSys IDREF #REQUIRED
            X_Coord CDATA      #IMPLIED
            Y_Coord CDATA      #IMPLIED
            Shape (CIRCLE|POLYGON|SQUARE) "CIRCLE">

<!-- the following "SimpleLink" points to the ID of a defined component
or functional_unit or cell_compartment or cellular_process.
The above can be accomplished by using IDREF instead of Simple Links.
However, it may be more extensible using links since we know that the
component definitions will be on the server somewhere (outside of any specific
xml doc) in the future. -->
<!ELEMENT ComponentLink (SimpleLink, Notes*)>
      <!ATTLIST ComponentLink
            NumberOfComponent CDATA #IMPLIED
            InCompartment IDREF #REQUIRED
            UniformInCompartment (TRUE|FALSE) "TRUE">

<!ELEMENT Synonym (Abbreviation*, Notes*)>
      <!ATTLIST Synonym
            Synonym CDATA #REQUIRED>
<!ELEMENT Abbreviation (#PCDATA|Notes)*>
      <!ATTLIST Abbreviation
            Abbreviation CDATA #REQUIRED>

<!-- having a RefLink element is for the sole purpose of making the
xml doc more readable; otherwise one would not know what the extended
link is all about since the "ExtendedLink" element is reused extensively.
In this case, the href attribute should point to some defined reference in
the same xml doc using XPointers: "#ID()"-->
<!ELEMENT RefLink (ExtendedLink)>

<!ELEMENT Ref ((Publication|Person|Organization)*, Notes*)>
      <!ATTLIST Ref
            Ref_ID ID #REQUIRED
            Date_Month CDATA #IMPLIED
            Date_Day CDATA #IMPLIED
            Date_Year CDATA #IMPLIED>

<!-- the following simplelink links to a medline record -->
<!ELEMENT Publication (Person*, SimpleLink, Note?)>
      <!ATTLIST Publication
```

```
                Journal CDATA #IMPLIED
                Publisher CDATA #IMPLIED
                PageStart CDATA #IMPLIED
                PageEnd CDATA #IMPLIED
                Volume CDATA #IMPLIED
                Issue CDATA #IMPLIED
                Type CDATA #IMPLIED
                Date_Month CDATA #IMPLIED
                Date_Day CDATA #IMPLIED
                Date_Year CDATA #IMPLIED>
<!ELEMENT Person (Organization*, Notes*)>
        <!ATTLIST Person
                FirstName CDATA #IMPLIED
                MiddleInit CDATA #IMPLIED
                LastName CDATA #IMPLIED
                StreetAddress CDATA #IMPLIED
                City CDATA #IMPLIED
                State CDATA #IMPLIED
                ZipCode CDATA #IMPLIED
                AreaCode CDATA #IMPLIED
                PhoneNum CDATA #IMPLIED
                Ext CDATA #IMPLIED
                Email CDATA #IMPLIED
                Web CDATA #IMPLIED
                Role CDATA #IMPLIED> <!-- "role" could be "contacting author" for
example -->
<!ELEMENT Organization (#PCDATA|Notes)*>
        <!ATTLIST Organization
                Name CDATA #REQUIRED
                Type (Commercial|Academic|Government) #REQUIRED>


<!-- "Role" describes the function of some item in a collection, such as "rate
limiting" -->
<!ELEMENT Transformations ((Transformation|Transformations|Effectors)*,
RefLink?, Notes*)>
        <!ATTLIST Transformations
                Transformations_ID ID #REQUIRED
                Transformations_Type CDATA #IMPLIED
                Transformations_Name CDATA #IMPLIED
                Role  CDATA #IMPLIED
                Group_Type CDATA #IMPLIED> <!-- such as coupled or simultaneous or
subpathway -->

<!ELEMENT Transformation (Input+,Output+,Effectors*,RefLink?, Notes*)>
        <!ATTLIST Transformation
                Transformation_ID ID #REQUIRED
                Transformation_Type CDATA #IMPLIED
                Transformation_Name CDATA #IMPLIED
                Role  CDATA #IMPLIED>

<!-- Input, Output and Effector reference Unit -->
<!ELEMENT Input (#PCDATA|Notes)*>
        <!ATTLIST Input
                Input_ID IDREF #REQUIRED>
```

```
        <!ATTLIST Output
                Output_ID IDREF #REQUIRED>
<!ELEMENT Effectors (Effector+, Notes*)>
        <!ATTLIST Effectors
                Group_Type (synergism|xyz) "synergism">
<!ELEMENT Effector (#PCDATA|Notes)*>
        <!ATTLIST Effector
                Effector_ID IDREF #REQUIRED
                Effect_Type CDATA #IMPLIED
                Role  CDATA #IMPLIED
                Is_Positive (TRUE|FALSE) "TRUE"> <!-- Effect_Type could be "enzyme"
-->


<!-- Feature-ID references an object of the type specified by Feature_Type -->
<!ELEMENT FeatureInfo (ExtendedLink, Notes*)>
        <!ATTLIST FeatureInfo
                Feature_ID          IDREF #REQUIRED
                Feature_Type        (Component|Unit|Transformations)     #REQUIRED
                Info_Type
(Entity|Assay|Compound|Reference|Pathway|Disease|Credibility) "Entity">

<!ELEMENT ExtendedLink (LinkLocator*, Notes*)>
        <!ATTLIST ExtendedLink
                XML-LINK CDATA #FIXED "EXTENDED"
                ROLE CDATA  #IMPLIED
                TITLE CDATA #IMPLIED
                INLINE (TRUE|FASLE)     "TRUE"
                SHOW (EMBED|REPLACE|NEW)      "REPLACE"
                ACTUATE (AUTO|USER)     "USER">

<!ELEMENT LinkLocator (#PCDATA|Notes)*>
        <!ATTLIST LinkLocator
                XML-LINK CDATA #FIXED "LOCATOR"
                ROLE CDATA  #IMPLIED
                HREF CDATA  #REQUIRED
                TITLE CDATA #IMPLIED
                SHOW (EMBED|REPLACE|NEW)      "REPLACE"
                ACTUATE (AUTO|USER)     "USER">

<!ELEMENT SimpleLink (#PCDATA|Notes)*>
        <!ATTLIST SimpleLink
                XML-LINK CDATA #FIXED "SIMPLE"
                HREF CDATA #REQUIRED
                TITLE CDATA #IMPLIED>

<!ELEMENT Notes (#PCDATA)>
```

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

(Attorney's Docket No. 99,068)

Applicant or
Patentee: Wang et al.

Serial or   Filed or
Patent No.   To be assigned   Issued: Herewith

Title: Database For Storage And Retrieval Of Cellular Information

VERIFIED STATEMENT CLAIMING SMALL ENTITY STATUS
(37 C.F.R. § 1.9(f) AND § 1.27(c)) - SMALL BUSINESS CONCERN

I hereby declare that I am

- [ ] the owner of the small business concern identified below:
- [X] an official of the small business concern empowered to act on behalf of the concern identified below:

NAME OF CONCERN   CELLOMICS, INC.

ADDRESS OF CONCERN   635 William Pitt Way, Pittsburgh, PA 15238

I hereby declare that the above-identified small business concern qualifies as a small business concern as defined in 13 C.F.R. § 121, and referenced in 37 C.F.R. § 1.9(d), for purposes of paying reduced fees to the United States Patent and Trademark Office, in that the number of employees of the concern, including those of its affiliates, does not exceed 500 persons. For purposes of this statement, (1) the number of employees of the business concern is the average over the previous fiscal year of the concern of the persons employed on a full-time, part-time, or temporary basis during each of the pay periods of the fiscal year, and (2) concerns are affiliates of each other when either, directly or indirectly, one concern controls or has the power to control the other, or a third party or parties controls or has the power to control both.

I hereby declare that rights under contract or law have been conveyed to and remain with the small business concern identified above with regard to the invention, entitled Database For Storage And Retrieval Of Cellular Information by inventor(s) Jian Wang, Chris Harrington, and Lans Taylor.

described in

- [X] the specification filed herewith.
- [ ] Application Serial No. _____ , filed _____ .
- [ ] Patent No. _____ , issued _____ .

- 1 -

If the rights held by the above identified small business concern are not exclusive, each individual concern or organization having rights in the invention must file verified statements averring to their status as small entities, and no rights to the invention are held by any person, other than the inventor, who would not qualify as an independent inventor under 37 CFR § 1.9(c) if that person made the invention, or by any concern which would not qualify as a small business concern under 37 CFR § 1.9(d), or a nonprofit organization under 37 CFR § 1.9(e).

Each person, concern or organization having any rights to the invention is listed below:

☒ No such person, concern or organization exists.
☐ Each such person, concern or organization is listed below.

Separate verified statements are required from each named person, concern or organization having rights in the invention averring to their status as small entities. (37 CFR § 1.27).

I acknowledge the duty to file, in this application or patent, notification of any change in status resulting in loss of entitlement to small entity status prior to paying, or at the time of paying, the earliest of the issue fee or any maintenance fee due after the date on which status as a small entity is no longer appropriate. (37 C.F.R. § 1.28(b))

I hereby declare that all statements made herein of my own knowledge are true and that all statements made on information and belief are believed to be true; and further that these statements were made with the knowledge that willful false statements and the like so made are punishable by fine or imprisonment, or both, under Section 1001 of Title 18 of the United States Code, and that such willful false statements may jeopardize the validity of the application, any patent issuing thereon, or any patent to which this verified statement is directed.

NAME OF PERSON SIGNING: LEE R. JOHNSTON JR.

TITLE IN ORGANIZATION: VICE PRESIDENT & CHIEF FINANCIAL OFFICER

ADDRESS OF PERSON SIGNING: 635 WILLIAM PITT WAY
PITTSBURGH PA 15238

Signature: _____
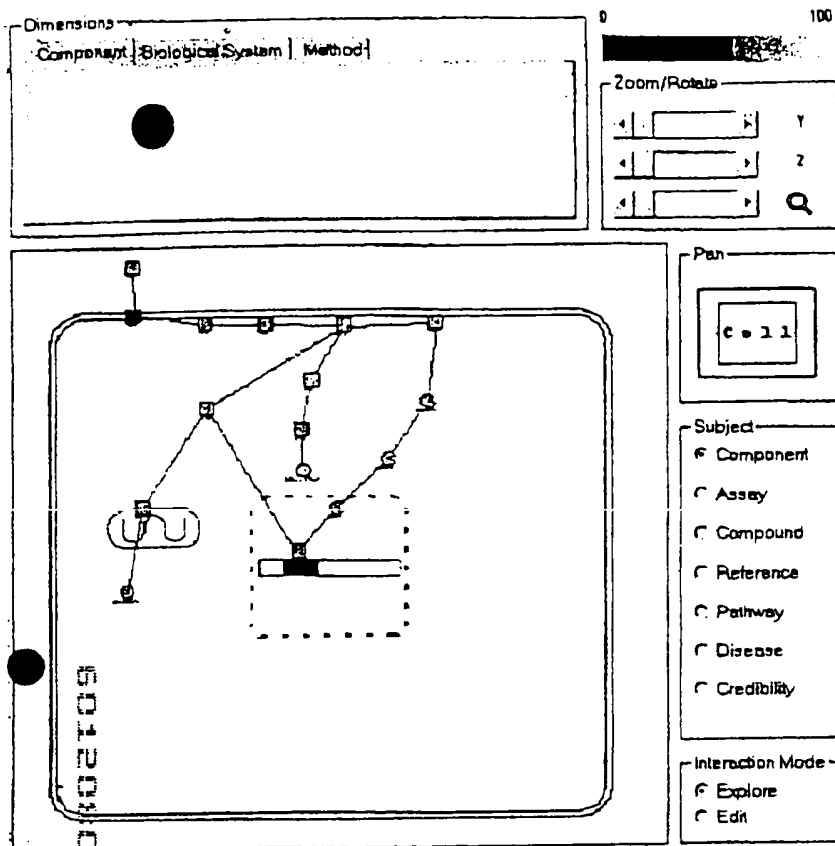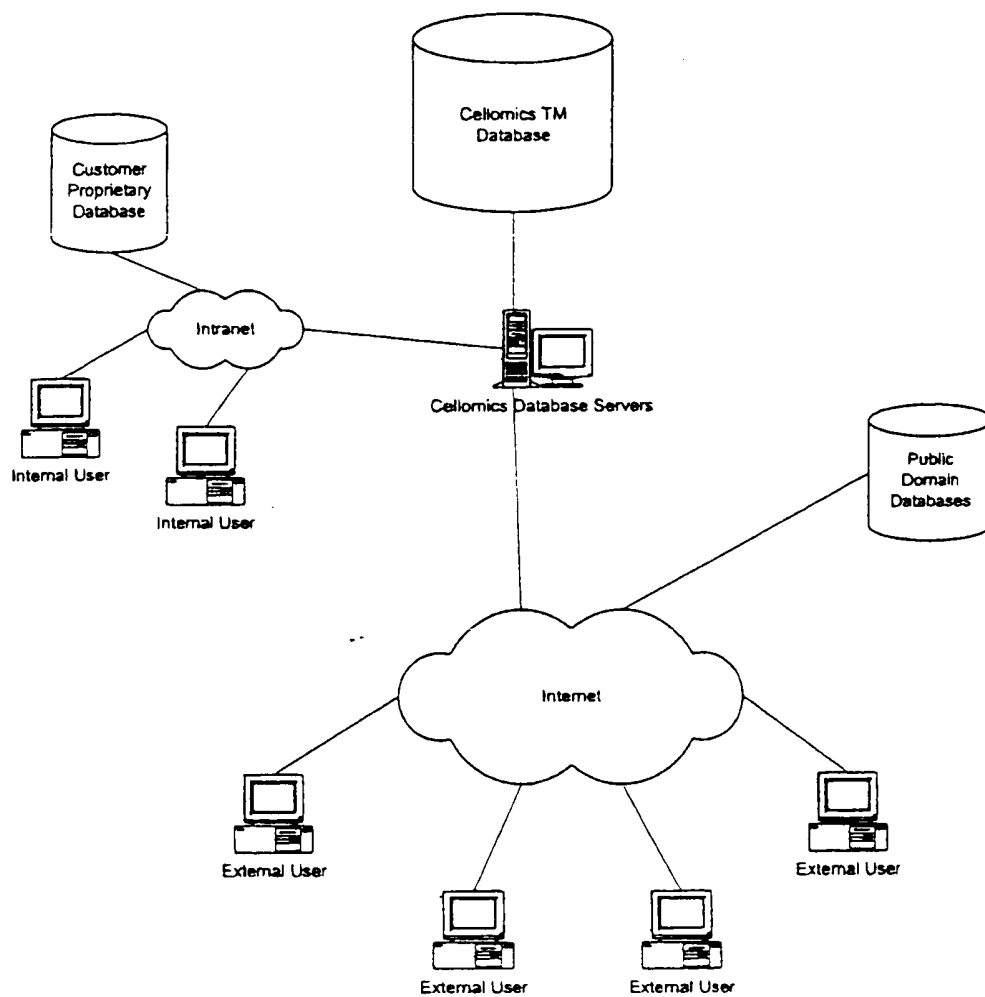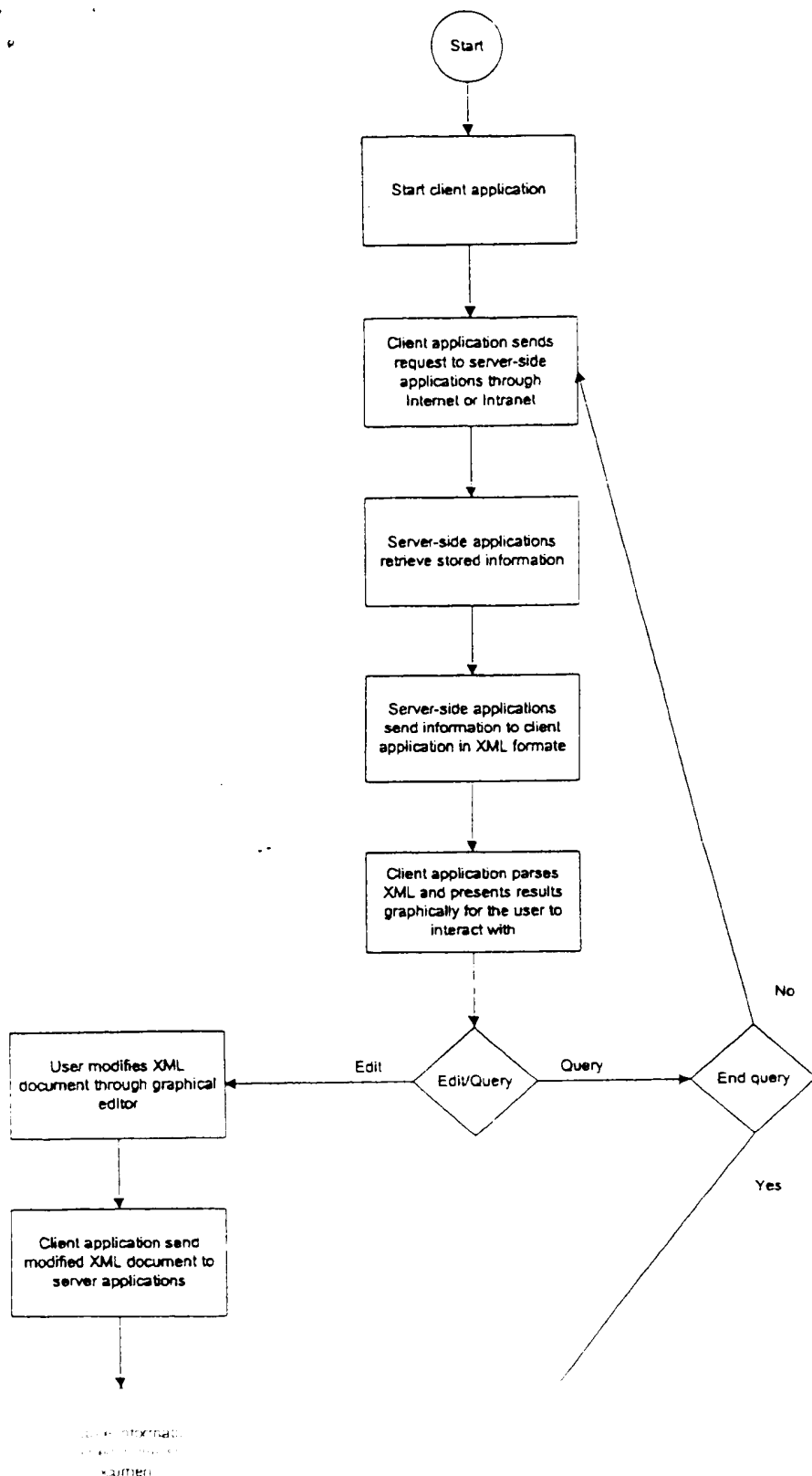
Date: 2/19/99

- 2 -

Fig 1

Customer
Proprietary
Database

Cellomics TM
Database

Intranet

Cellomics Database Servers

Internal User

Internal User

Public
Domain
Databases

Internet

External User

External User

External User

External User

```
                              ( Start )
                                 │
                                 ▼
                  ┌────────────────────────────┐
                  │   Start client application  │
                  └────────────────────────────┘
                                 │
                                 ▼
                  ┌────────────────────────────┐
                  │  Client application sends   │
                  │  request to server-side     │
                  │  applications through       │
                  │  Internet or Intranet       │
                  └────────────────────────────┘
                                 │
                                 ▼
                  ┌────────────────────────────┐
                  │  Server-side applications   │
                  │  retrieve stored information │
                  └────────────────────────────┘
                                 │
                                 ▼
                  ┌────────────────────────────┐
                  │  Server-side applications   │
                  │  send information to client │
                  │  application in XML formate │
                  └────────────────────────────┘
                                 │
                                 ▼
                  ┌────────────────────────────┐
                  │  Client application parses  │
                  │  XML and presents results   │
                  │  graphically for the user to │
                  │  interact with              │
                  └────────────────────────────┘
                                 │
                                 ▼
```

┌────────────────────┐   Edit   ◇ Edit/Query ◇   Query   ◇ End query ◇   No
│ User modifies XML │◄─────────   ─────────►
│ document through   │
│ graphical editor   │   Yes
└────────────────────┘
         │
         ▼
┌────────────────────┐
│ Client application send │
│ modified XML document to │
│ server applications │
└────────────────────┘
         │
         ▼
   [illegible text]